Speech Processing 2009/10

3rd Test

June 7th 2010

Please identify this form with your name and student number in the reserved space at the bottom. The answers to multiple-choice questions will only be accepted if inserted in the appropriate place. Wrong answers to these questions will be penalized. The phonetic symbols should use the SAMPA alphabet (Lisbon accent).

1. Classify as True (T) or False (F)
   (a) In keyword spotting systems, a high percentage of false alarms is typically associated with the selection of a high threshold.
   (b) Cepstral derivatives model the perceptually relevant aspects of temporal changes in the spectrum.
   (c) The coda part of the syllable is the one which is most often pronounced according to the canonical pronunciation.
   (d) Distortion measures such as the Itakura-Saito distance explore the spectral envelope parameterization achieved by LPC techniques.
   (e) Error bursts are often caused by OOV words.
   (f) Unsupervised speaker adaptation requires a large amount of orthographicall labeled speech data.
   (g) Language model interpolation is frequently used to combine the LM build with a large amount of in-domain data with the LM build with a small amount of out-of-domain data.
   (h) Forced alignment may be used to get estimates of phone and word boundaries.
   (i) When the amount of training data is small, high order n-gram models are advisable.
   (j) Noisy environments frequently lead to situations when no clear search path dominates.
2. Give examples (one per class) of
   (a) Vector quantization technique
   (b) An acoustic parametrization technique different from MFCC
   (c) An alternative to discrete distribution or Gaussian mixture distributions
   (d) HMM decoding algorithm
3. Which models would you choose to use in the first stage of a multi-pass decoding system?
   (a) bigram or quadrigram language models?
   (b) context dependent or context independent phone models?
4. Single-channel enhancement methods such as spectral subtraction (list all that apply)
   (a) are designed for stationary noise environments
   (b) show a better performance than multi-channel enhancement methods
   (c) perform train-test mismatch adaptation in the feature space
5. What are the axes in ROC plots of the performance of keyword spotting systems?
6. In a demo of the McGurk effect, a recording of the syllable *ba* is shown simultaneously with the viseme of the syllable *ga*. What is the syllable that most speakers will hear?

7. A WER of 40% happens frequently in the following scenarios (list all that apply)
   (a) SI connected digit recognition (unknown string length)
   (b) SI large-vocabulary spontaneous speech recognition in clean conditions
   (c) SI large-vocabulary read speech in clean conditions
   (d) SI large-vocabulary read speech in very low SNR conditions
   (e) SI spontaneous speech in clean conditions, small perplexity task
   (f) SD large-vocabulary read speech in clean conditions
8. Name the five main blocks of a large vocabulary speech recognition system.
9. The following extract was produced by a speaker independent large vocabulary continuous speech recognition system for a TED Talk by Carolyn Porco.

   Saturn is accompanied by a very large and diverse collection of moon they range in size much you call it is across, to you is to get caught in the US. Know several beautiful pictures were taken of Saturn, in fact to show said earning accompaniment with some of its most

   The corresponding manual transcription is below:

   Saturn is accompanied by a very large and diverse collection of moons. They range in size from a few kilometers across, to as big across as the US. Most of the beautiful pictures we've taken of Saturn, in fact, show Saturn in accompaniment with some of its moons.

   Ignoring punctuation and capitalization, compute the corresponding values of H ("correct"), D ("deletions"), S ("substitutions"), I ("insertions"), N ("total"), %Corr, %Acc and %WER. Compute as well the number of deletions of full stops.
10. Consider the training corpus that consists of the following sentences:
    The cat is behind the tree.
    The chair is below the window.
    The orange chair is very small.
    The tree is very tall.
    The orange cat is on the big chair.

    Consider the test sentence:
    The big cat is behind the window.
    (a) Compute the number of unigrams, bigrams and trigrams of the training corpus (excluding ¡/s¿ ¡s¿ transitions), and the dimension of the vocabulary.
    (b) Compute the probability of the test utterance using a bigram language model without any type of smoothing.
    (c) Compute the probability of the test utterance using a bigram language model with add-one smoothing.
    (d) Build a test sentence as long as possible with a non-zero probability according to a trigram model.
    (e) What is the most frequent trigramin the training corpus?

Test 3 - Answers

Name:	
Number:	

1. (3.0 val.) Indicate T or F:

a	b	c	d	e	f	g	h	i	j

2. (2.4 val.)

a	
b	
c	
d	

3. (1.0 val.)

a	
b	

4 to 7 (0.9/1.0/1.0/1.0 val.)

4	
5	
6	
7	

8. (2.0 val.)

9. (3.5 val.)

H	D	S	I	N	% Corr	% Acc	% WER	Del-Stop

10. (4.2 val.)

a)	
b)	
c)	
d)	
e)	